

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-232934

(43)Date of publication of application : 02.09.1998

(51)Int.Cl.

G06T 7/00

(21)Application number : 09-033704

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 18.02.1997

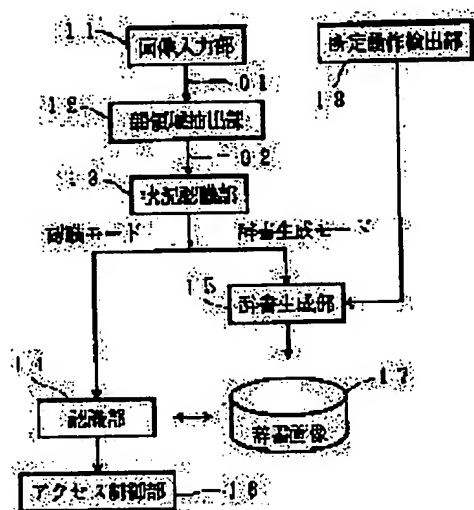
(72)Inventor : FUKUI KAZUHIRO  
YAMAGUCHI OSAMU

## (54) FACE IMAGE REGISTERING DEVICE AND ITS METHOD

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide the face image registering device which can exclude the registration of an unnecessary face image without placing any burden on a user and also increase dictionary precision.

**SOLUTION:** An information terminal device 10 with security consists of an image input part 11, a face area extraction part 12, a state recognition part 13, a recognition part 14, a dictionary generation part 15, an access control part 16, dictionary images 17, and a specific operation detection part 18. After the specific operation detection part 18 detects specific operation of the user, the face area extraction part 12 extracts an image of a face area from an image inputted from the image input part 11 and a dictionary generation part 15 judges whether or not the image is valuable enough to register as a dictionary image 17.



## LEGAL STATUS

[Date of request for examination] 13.09.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3469031

[Date of registration] 05.09.2003

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

## CLAIMS

## [Claim(s)]

[Claim 1] In the face image registration equipment which has a dictionary means by which a person's face image is registered A face field extract means to extract the input face image which is an image of said person's face field out of the image inputted by image input means to input said person's image, and said image input means, A decision means to judge whether the input face image extracted by said face field extract means is a face image which can be registered into said dictionary means, Face image registration equipment characterized by consisting of a registration means made to register into said dictionary means with said decision means by using as said face image the input face image judged that registration is possible.

[Claim 2] Said decision means is face image registration equipment according to claim 1 characterized by extracting the characteristic quantity of a face from the input face image extracted by said face field extract means, comparing this extracted characteristic quantity with the criteria set up beforehand, and judging that registration is possible if said characteristic quantity is said beyond criteria.

[Claim 3] It is face image-registration equipment according to claim 1 characterized by to judge whether it has a detection means of operation detect predetermined actuation of said person, and said decision means can register as said face image for the input face image extracted by said face field extract means between fixed periods after the detection while said detection means of operation has detected said predetermined actuation to register with said dictionary means.

[Claim 4] The image input step which inputs said person's image in the face image registration approach for registering a person's face image, The face field extract step which extracts the input face image which is an image of said person's face field out of the image inputted in said image input step, The face image registration approach characterized by the input face image extracted in said face field extract step consisting of a decision step which judges whether it is the face image in which said registration is possible, and a registration step into which the input face image judged that registration is possible in said decision step is made to register as said face image.

[Claim 5] Said decision step is the face image registration approach according to claim 4 characterized by extracting the characteristic quantity of a face from the input face image extracted in said face field extract step, comparing this extracted characteristic quantity with the criteria set up beforehand, and judging that registration is possible if said characteristic quantity is said beyond criteria.

[Claim 6] It is the face image-registration approach according to claim 4 characterized by to judge whether it has the detection step of operation which detects predetermined actuation of said person, and the input face image extracted in said face field extract step between the fixed period after detection can register said decision step as said face image while having detected said predetermined actuation in said detection step of operation.

[Translation done.]

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

## DETAILED DESCRIPTION

---

### [Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the face image registration equipment which extracts a face field out of the inputted image, and carries out dictionary registration of the face automatically, or its approach.

[0002]

[Description of the Prior Art] A face image recognition technique is a component engineering indispensable to construction of a human interface or a security system. Face image recognition consists of a face field extract, a face focus extract, and discernment. About the latest technical trend, reference [latest research trend [ of a Shiono \*\*\*\* Sanada / Hidehiko / " personal authentication technique ]", Shingaku Giho OSF92-17.] is detailed. The conventional face image identifying method can be classified into two of the roughly divided followings.

[0003] One is the approach of calculating similarity with a candidate person's feature vector which parameterizes the location of the focus, such as an eye, a nose, and opening, a configuration, and size, generates a feature vector, and is registered beforehand. The person showing a dictionary vector with the highest similarity is discriminated from a person in question. These are classified into the structural-analysis-technique.

[0004] Another approach is an approach based on the similarity of the pattern of the image by which a location and size were normalized by geometrical conversion of the two-dimensional affine conversion on the basis of the focus, such as a pupil and a nose, etc., and the dictionary normalization image registered beforehand. The person who expresses a dictionary image with the highest similarity like the former is discriminated from a person in question. These can be classified into the technique like a pattern according to the approach put in practical use in the conventional character recognition.

[0005] In both [ of the two above-mentioned technique ] cases, statistical distance, such as a correlation value between images and Euclidean distance in the inside of a feature space, is used as similarity. Approaches, such as various pattern recognition theory put in practical use in character recognition, for example, a subspace method, [the Hidemitsu Ogawa written by ERUKKI OYA and Sato \*\*\*\* and "pattern recognition and subspace-method" Sangyo Tosho Publishing (1986)], and compound similarity, are applicable to this discernment processing. The configuration of the identifying method is detailed to [Funakubo \*\* "pattern recognition" KYORITSU SHUPPAN (1991)], [Taizo Iijima "pattern recognition theoretical" Morikita Shuppan (1989) etc.], etc.

[0006] In addition to the above-mentioned face recognition processing, on the occasion of actual application of the above-mentioned face recognition, registration of a user, i.e., a registrant's face dictionary generation, becomes indispensable. For example, in the handwriting recognition of Chinese character using a subspace method, the dictionary was generated from the study sample image of hundreds of sheets for every alphabetic character. As for a three-dimension body like the target face, compared with an alphabetic character, large still a lot of [ fluctuation of configurations, such as lighting conditions, face sense, and expression change, or brightness ] study sample images are needed by this invention [Murase \*\*, sherry NAIYA, "the three-dimension object recognition by two-dimensional collating", IEICE TRANSACTIONS (D-II) J77-D-II, 11, pp.2179-2187, and 1994.].

[0007] The conventional dictionary generation Reference [red Matsushige, Sasaki \*\*, Akio Fukamachi, Yasuhito Suenaga, "the robust method of identifying a transverse-plane face by shade image matching", IEICE TRANSACTIONS (D-II), J76-DII, 7, pp.1363-1373, 1993.], [Yoshie Komatsu, Yasuo Arika, "logging recognition of the face by the sense using a subspace method", PRU 95-191, pp.7-14, 1996.], [-- M.Turk and A.P.Pentland : [ ] -- "Face recognition using eigenfaces" -- Proc.CVPR 11, pp.453-458, and 1993. --] -- [Alex Pentland and Baback Moghaddam and Thad Starner, So that "View-based and modular eigenspaces for face recognition", CVPR'94, pp.84-91, 1994.], etc. may see The help was performing logging of a face and selection of an image from the image which turned and photoed

a registrant's face under good lighting conditions in the direction of predetermined, for example, a transverse plane, 15 degrees of right and left, and 15 degree \*\* of vertical.

[0008] A physiological change of the effect of the condition of the day, expression change and hair, and a mustache etc. tends to produce a face. Therefore, in order to realize a high recognition rate, it is necessary to update a face dictionary for every fixed period. However, the frequent renewal of a dictionary will raise a user's burden. So, in order to make face recognition easy to use, it becomes important how a burden is mitigated and registered without making a user conscious.

[0009] For example, when it is interrupted temporarily and leaves the activity in an information terminal, the function to prevent others' access is considered. although using the dictionary generated beforehand is also considered -- him -- it is more effective to use the dictionary image generated as much as possible immediately before, in order to gather identification and the rate of others exclusion. However, in order to make a user's burden high, dictionary registration of carrying out dictionary registration at every \*\*\*\* is carried out without a user's making it conscious while accessing information machines and equipment.

[0010] However, the study sample collected without making a registrant conscious contains the data which are not effective as study data the case where an eye is usually closed, when expression is changing a lot. If dictionary generation is performed from the study sample in which these error images are contained, the precision of a dictionary will fall. In order to raise dictionary precision, it is necessary to increase the number of study data, and real-time dictionary generation cannot be coped with in selection by the help.

[0011]

[Problem(s) to be Solved by the Invention] Conventionally, the help was performing generation of a face dictionary using predetermined, for example, a transverse plane, 15 degrees of right and left, and the face image that was made to turn a registrant's face in the direction for 15 degrees of every upper and lower sides, and was photoed in advance under good lighting conditions. A face changes with time amount. Therefore, in order to realize a high recognition rate, it is necessary to update a face dictionary for every fixed period. However, performing frequent renewal of a dictionary will raise a user's burden.

[0012] About the automatic study sample collection for dictionary generation, if an approach [ finishing / application / already ] [Japanese Patent Application No. No. 61463 [ eight to ]] is applied, it is realizable. However, the technical problem that much data which are not effective in dictionary generation are contained remains in the study data collected without still making a registrant-ed conscious. In order to raise dictionary precision, it is necessary to increase the number of study data, and real-time dictionary generation is difficult to realize in selection by the help.

[0013] Then, this invention provides a user with a burden, \*\*\*\*, the face image registration equipment that can eliminate registration of an unnecessary face image and can raise dictionary precision, and its approach.

[0014]

[Means for Solving the Problem] In the face image registration equipment which has a dictionary means by which this invention has registered a person's face image A face field extract means to extract the input face image which is an image of said person's face field out of the image inputted by image input means to input said person's image, and said image input means, A decision means to judge whether the input face image extracted by said face field extract means is a face image which can be registered into said dictionary means, It is face image registration equipment characterized by consisting of a registration means made to register into said dictionary means with said decision means by using as said face image the input face image judged that registration is possible.

[0015] Moreover, this invention is set to the face image registration approach for registering a person's face image. The face field extract step which extracts the input face image which is an image of said person's face field out of the image inputted in the image input step which inputs said person's image, and said image input step, The decision step the input face image extracted in said face field extract step judges it to be whether it is the face image in which said registration is possible, It is the face image registration approach characterized by consisting of a registration step into which the input face image judged that registration is possible in said decision step is made to register as said face image.

[0016] It judges whether it is effective as study data by comparing the criteria beforehand set up in the characteristic quantity which can be found from the face field extracted, for example as it is the above-mentioned invention. The image which normalized as characteristic quantity on the basis of the magnitude of the face focus, such as a pupil, a nostril, and \*\*\*\*, the amounts which can be found from positional information, or these points is used. It becomes possible to remove automatically the image from which expression changed a lot by this, the image which closed the eye, the image which opened opening from a study image.

[0017] Furthermore predetermined actuation of mouse actuation, keyboard grabbing, button grabbing, etc. is detected, and dictionary generation is performed only for the image inputted during actuation. It decreases [ the time amount of

useless study data collection ] for being aimed only at under actuation and is efficient. Moreover, possibility of collecting the images which are not effective as dictionaries in the case of having turned to width etc. also decreases. According to this invention, the equipment and the approach of extracting a face field out of the inputted image, and carrying out dictionary registration automatically are realizable.

[0018]

[Embodiment of the Invention] Hereafter, one example of this invention is explained based on drawing 7 from drawing 1.

[0019] In information terminal units, such as a personal computer, in case under an activity is interrupted temporarily and a terminal is left, it is a required function to prevent others' access. Therefore, this example explains taking the case of the information-machines-and-equipment terminal unit 10 which carried the function to prevent others' access at the time of \*\*\*\*.

[0020] Drawing 1 is the block diagram showing the outline of the information terminal unit 10 with security concerning this example.

[0021] This equipment 10 consists of the image input section 11, the face field extract section 12, the situation-recognition section 13, a private seal 114, the dictionary generation section 15, the access-control section 16, a dictionary image 17, and a predetermined actuation detecting element 18.

[0022] (Image input section 11) The image input section 11 is for inputting the image of the person who becomes a candidate for recognition, for example, consists of a TV camera. The image 01 inputted from this image input section 11 is digitized by the A/D converter, and is sent to the face field extract section 12. For example, a TV camera is installed in the lower part of the monitor of the information terminal unit 10. Or you may install in the rectangular head of a monitor.

[0023] (Face field extract section 12) The face field extract section 12 always continues extracting the face field image 02 from the input image sent from the image input section 11. while moving the standard face image (template) registered beforehand over a full screen in this example -- a correlation value -- calculating -- the highest correlation value -- \*\*\*\* -- let the field which is be a face field. Let the partial maximum point of a correlation value be a face field candidate. Suppose that a face does not exist, when lower than the threshold to which the correlation value in this face field candidate was set. It is set as 30, specifically using the maximum correlation value as 100 as a threshold. Since it corresponds to sense change of a face, if two or more templates are used by compound similarity etc., a face field can be further extracted to stability. This processing may be transposed to the extraction method based on the color information described previously.

[0024] (Situation-recognition section 13) In the situation-recognition section 13, a user's condition (under an activity, \*\*\*\*, taking a seat) is discriminated from a time change of a face field extract result, and the change in dictionary generation mode and recognition mode is performed.

[0025] It is set as recognition mode at the periods from T0 to T1 the time of a face field beginning to be detected as shown in drawing 2. the recognition processing described in this mode later -- him -- identification processing is performed. It becomes accessible when identified him at the T1 time.

[0026] When it changes to dictionary generation mode after T1 time and the dictionary generation signal from the predetermined actuation detecting element 18 is inputted, collection of study data is performed. It judges that it \*\*\*\*(ed) when it became impossible to have detected a fixed period face field, and a screen lock is started immediately. Or a screen lock may be started, after conveying the main point to a user with light, voice, etc. and seeing a user's reaction. Time of day T1 can be changed according to an object model or the class of access actuation here.

[0027] (Recognition section 14) The recognition section 14 consists of face focus extract section 14a, normalization image generation section 14b, and pattern-matching section 14c. The block diagram of the recognition section 14 is shown in drawing 3.

[0028] In face focus extract section 14a, the focus, such as a pupil, a nose, and \*\*\*\*, is extracted from the inside of the extracted face field. The approach [Japanese Patent Application No. No. 61463 [ eight to ]] which combined the configuration information for which it has already applied, and pattern information is applicable.

[0029] The fundamental idea of this approach asks for the candidate of the focus using the configuration information that location precision is high, and verifies it by pattern matching. Since this approach positions using configuration information, it can expect a high location precision. Moreover, since matching which used the multi-template for selection of the right focus from a candidate group is applied, it is robust to fluctuation of the configuration brightness of the focus. Since pattern matching is carried out about processing speed only to the candidate who narrowed down with the degree-of-separation filter with little count cost, compared with the approach of carrying out pattern matching of the whole, drastic reduction of computational complexity is realizable.

[0030] In addition to this Approach [A.L.Yuille based on edge information, "Feature extraction from faces using deformable templates" and IJCV, vol.8 : 2 and pp.99-111, 1992.][Shizuo Sakamoto, the Miyao proton, Joji Tajima, "a focus extract of the eye from a face image", IEICE TRANSACTIONS D-II, Vol.J76-D-II, No.8, pp.1796-1804, August, 1993.], and a proper space method Applied Eigen feature Method [Alex Pentland, Baback Moghaddam, Thad Starner, and "View-based and modular eigenspaces for face recognition", The approach [the Sasaki \*\*\*\* Akamatsu \*\*, Yasuhito Suenaga, "the method of alignment of the face using the color information for face image recognition", IE 91-2, pp.9-15, and 1991.] based on CVPR'94, pp.84-91, 1994.], and color information is applicable.

[0031] In normalization image generation section 14b, it normalizes on the basis of the focus. The example of the normalization processing on the basis of a pupil and a nostril is shown in drawing 4 . The sense of vector E1E2 is amended in parallel, further, from on c1 and c2, the main coordinate of a normalization image and breadth are set up by 2 twice vector E1E2, and a dip sets up the point CP of 1/3 by 2 twice vector c1c2.

[0032] In pattern-matching section 14c, it asks for pattern similarity as compared with a normalization image and the face image currently stored in the dictionary image. It identifies being him, when pattern similarity is higher than a reference value. In being smaller than criteria, it considers as others. As the approach of pattern matching, a correlation technique, a subspace method, a compound similarity method, etc. are applicable.

[0033] (Predetermined actuation detecting element 18) The predetermined actuation detecting element 18 detects whether the user is performing predetermined actuation of a keyboard, a carbon button push, or actuation of a mouse. The dictionary generation section 15 is controlled based on a detection result.

[0034] Since it is suitable to specifically carry out dictionary registration of the user performing predetermined actuation of keyboard grabbing, a carbon button push, actuation of a mouse, etc. since possibility that the transverse-plane image of a user's face is reflected to the image input means 11 is high when it detects, while performing predetermined actuation, a dictionary generation signal is transmitted to the dictionary generation section 16.

[0035] (Dictionary generation section 16) The dictionary generation section 16 consists of face focus extract section 14a, frame evaluation section 16a, normalization image generation section 14b, study image storage section 16b, and principal-component-analysis section 16c. With the recognition section 14, face focus extract section 14a and normalization image generation section 14b are common, and are used here. The block diagram of the dictionary generation section 16 is shown in drawing 5 .

[0036] If a dictionary generation signal is inputted from the predetermined actuation detecting element 18, as for dictionary generation, collection of study data will be performed.

[0037] In frame evaluation section 16b, it judges whether the face field image 02 extracted from the information on the extracted face focus is effective in dictionary generation. Although three valuation bases are applied here, you may apply in individual or combination.

[0038] (i) The physical relationship of each focus is specified beforehand, and when this physical relationship collapses, it is judged that it is not effective as study data. For example, they are both pupils, a nostril, the physical relationship of \*\*\*\*, etc.

[0039] (ii) The proper face image and pattern similarity which were stored in the dictionary image 17 are made evaluation. When similarity is smaller than criteria, it is judged that it is not effective. The homogeneity of study data will be raised, so that criteria are made high. When too high, it becomes impossible however, to correspond to change of the face sense etc.

[0040] Or the dictionary image generated from the study image normalized by the right focus in advance instead of the proper face image stored in the dictionary image 17 may be used.

[0041] If it furthermore takes into consideration to the face sense, it is possible to remove the face which turned to width extremely. In this case, when the comparison with the face dictionary pattern currently beforehand prepared for every face sense of a direction 1 - N is performed and it is in agreement with the dictionary of the predetermined face sense, it is judged that it is effective in dictionary generation.

[0042] (iii) The study data which investigated the relation of the focus extracted for every frame, and closed the eye are detected. The coordinate of the focus extracted newly can be expressed by the linear combination of the coordinate of the focus already extracted with the frames 1-4 before that. Therefore, if the focus is correctly extracted with all frames, the error at the time of carrying out linear combination will become small. Conversely, an error becomes large, when an eye is closed and eyebrows and a pupil are mistaken.

[0043] Drawing 6 is used and explained to a detail about this.

[0044] If an orthographic projection model is assumed, the two-dimensional coordinate of the face focus seen from the direction of arbitration, without having a three-dimension model By the linear combination of the two-dimensional coordinate of the focus matched on the image of four sheets as shown in a formula (1) and (2) [S.Ullman which can be

expressed, and R. Basri: "Recognition by Linear Combinations of Models, IEEE Trans.PAMI, Vol.13, No.10, pp.992-1006, 1991.][Yasuhiro Mukogawa, Yuichi Nakamura, Tomoichi Ota, "generation of the face image of the direction of arbitration using two photographs of his face", \*\*\*\*\*, Vol.37, and No. -- 4, pp.635-644, and 1996. --]. The two-dimensional coordinate value (X, Y) of the focus newly detected from two-dimensional coordinate value [ of the right focus in four frames already extracted in the dynamic-image train using this property ] (x1 and y1) - (x4 and y4) is verified.

[0045] It asks for linear combination multiplier  $a_i$ ,  $b_i$ , ( $i=1$ , and 4) with a least square method from the two-dimensional coordinate of the focus to the image of four frames, and a new coordinate (X, Y). The error Res which calculates (x', y') for an approximation coordinate from the coupling coefficient obtained conversely (X, Y) is calculated shortly.

[0046]

[Equation 1]

$$X = a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot x_4 \quad (1)$$

$$Y = b_1 \cdot y_1 + b_2 \cdot y_2 + b_3 \cdot y_3 + b_4 \cdot y_4 \quad (2)$$

$$Res = \sqrt{(X - x')^2 + (Y - y')^2} \quad (3)$$

Error Res shows the coordinate error of the new focus. Since the error has arisen in the extract location when an error is larger than a threshold, this data is not effective as study data.

[0047] Above (i) - (iii) It is also possible to judge by whether distribution of the average luminance of a face field or brightness is contained in the criteria range besides the valuation basis.

[0048] study image storage section 16b -- the above -- the study data judged to be effective are stored. If the stored number of study data reaches the number of regular, a study image will be sent to principal-component-analysis section 16c.

[0049] In principal-component-analysis section 16c, principal-component-analysis (KL expansion) application is carried out to the image data stored in study image storage section 16b, and it asks for a proper image. The proper image of a high order N individual is stored in the dictionary image 17 from the one where characteristic value is larger.

[0050] Although the above-mentioned example explained the example which performs dictionary generation on real time from a dynamic image, it is also possible to carry out, once it stores in memory only the frame by which the detecting signal from a predetermined actuation detecting element came from the dynamic image. In this case, since possibility effective in dictionary generation is aimed only at a high frame, improvement in the speed and stability of processing can be measured.

[0051] Moreover, although the subspace method which is the technique like a pattern is applied to discernment in the above-mentioned example, you may transpose to structural technique.

[0052] (Explanation of actuation of this equipment 10) Next, actuation of this example is explained along with drawing 7.

[0053] The first access performs the usual log in which enters a password etc. Or it is also possible to perform an automatic log in using the dictionary of the user who registered in advance.

[0054] In the face field extract section 12, a face field is extracted from the image first inputted from the image input section 11, and it is sent to the situation-recognition section 13. Moreover, in the predetermined actuation detecting element 18, predetermined actuation detecting signals, such as mouse actuation, keyboard grabbing, and button grabbing, are sent to the situation-recognition section. In the situation-recognition section 13, the change in recognition mode and dictionary generation mode is performed based on the time existence of a face field. Moreover, it controls whether study data are collected based on a predetermined actuation detecting signal.

[0055] time of day T0 to T1 sets it as recognition mode first -- having -- the last user -- it identifies whether you are him.

[0056] (i) A recognition mode face field is sent to face focus extract section 14a, and a pupil, a nostril, and \*\*\*\* are detected. Next, normalization is performed on the basis of these focus. Similarity with a user's dictionary image is calculated last time by pattern-matching section 14c, and when similarity is higher than a reference value, a normalization image is identified him who was using it last time, and becomes accessible. Access is forbidden when smaller than a reference value.

[0057] Next, from time of day T1, it is set as dictionary generation mode.

[0058] (ii) When the dictionary generation mode predetermined actuation detecting element 18 detects predetermined actuation and the predetermined actuation detecting signal is turned on, study data are collected and a dictionary image is generated. The face field image extracted from the image inputted continuously is sent to face focus extract section

14a, and a pupil, a nostril, and \*\*\*\* are detected. The extracted focus information is sent to frame evaluation section 16a, for example, an error is calculated from a formula (3). When an error is smaller than a threshold, it is judged that it is effective as study data, and it is stored in study image storage section 16b. If the study data more than convention number of sheets are stored in study image storage section 16b, study data will be sent to principal-component-analysis section 16c, and a dictionary image will be generated.

[0059] If a user \*\*\*\*, a screen lock will be started immediately. If a user takes a seat again, it will return to the first processing.

[0060] In addition, in the above-mentioned explanation, when the predetermined actuation detecting element 18 detected predetermined actuation and the predetermined actuation detecting signal was turned on, the dictionary image was generated, but further, after a predetermined actuation detecting signal suits at OFF, a dictionary image may be continuously generated between fixed periods (for example, for 1 minute).

[0061] (Example of modification) Next, the example in service of not only the security of an information terminal but an information terminal is explained.

[0062] For example, when registering a face, it considers memorizing to coincidence not only the characteristic quantity obtained from a face field but the actuation hysteresis of an information terminal. while operating discernment of the human being, simultaneously a terminal the information on a face, simultaneously by registering, it can extract from hysteresis what kind of information was searched for.

[0063] If information retrieval is made into an example, a user will do the activity which inputs a retrieval type etc. In addition to information required for identification, the hysteresis of the retrieval type and the contents of retrieval is stored in the database. Based on the stored data, the user's interesting contents are presumed and extracted and it holds by performing new information retrieval.

[0064] And again, when a user performs a log in to an information terminal, and an actuation return, the service of him which offers automatically information considered to be the request newly searched to the terminal is attained.

[0065] By linking such hysteresis information with the registration information on a face, and being stored, it may become instead of the time stump when to have retrieved such information, it may recognize whether it is close to the condition of the face at the time of when based on this, and the information based on the hysteresis information on of those days may be shown.

[0066] Moreover, as a verification device in the case of registration of a face, if expression detection and recognition are used, it is also possible to take the actuation hysteresis of the information according to human being's feeling condition, and the dictionary at the time of usual, the dictionary when laughing, the dictionary when being angry, etc. can also be generated. In case face recognition is newly performed, information offer corresponding to expression information may be performed.

[0067] When these perform continuously renewal of face registration, and renewal of information actuation hysteresis to coincidence, clearer information offer is attained.

[0068] Although these examples explained the information terminal to the example, they may be transposed, for example to home electronics, such as TV and a telephone.

[0069] In the case of TV, a subminiature camera is carried in remote control. Predetermined actuation is defined as the actuation which is pushing the carbon button of remote control. The actuation which in the case of a telephone micro CCD carries in an earphone and raises an earphone is defined as predetermined actuation.

[0070] Furthermore, the equipment 10 of this example may be used for an automatic teller's machine (henceforth ATM).

[0071] namely, him, a user,, by using this equipment 10 for a check, it forgets be based on others and the unauthorized use of usurpation of cash etc. can be prevented certainly. In this case, the image input section 11 is arranged near the control panel of ATM, and it is made for a user's face to be reflected certainly. Moreover, the account number and updating registration time of day may be registered into a face image and coincidence.

[0072]

[Effect of the Invention] As mentioned above, a new face image is extracted out of the image which was inputted according to this invention, in order to register only the input face image which judged whether it was the face image with which this deserves dictionary registration, and was judged to be effective, registration of an unnecessary face image can be eliminated and dictionary precision can be raised.

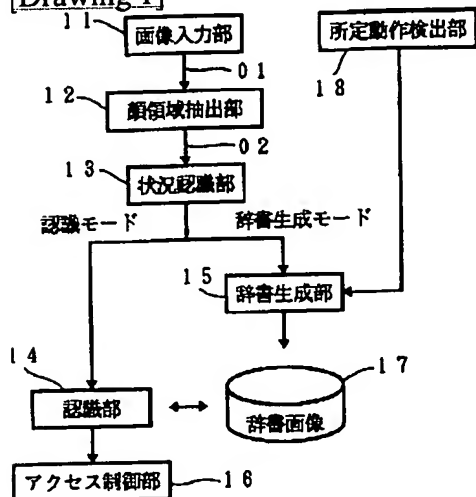
## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

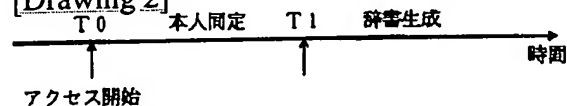
1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

## DRAWINGS

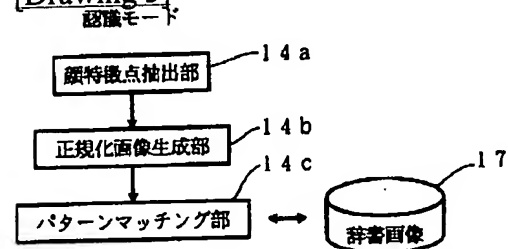
[Drawing 1]



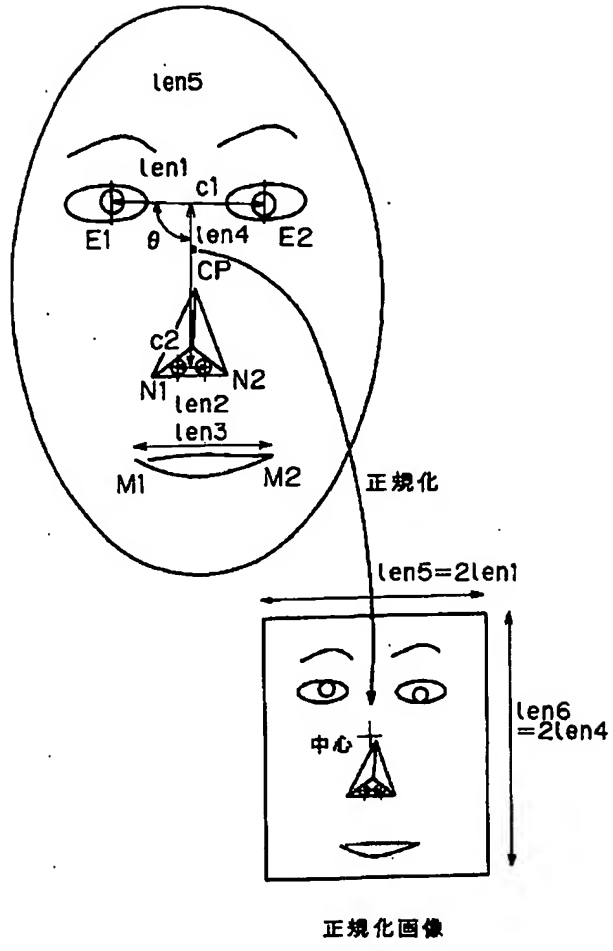
[Drawing 2]



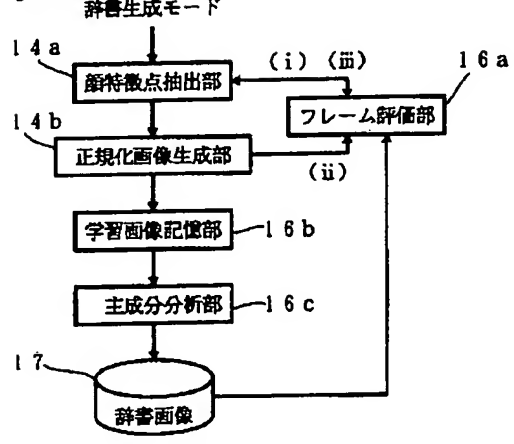
[Drawing 3]



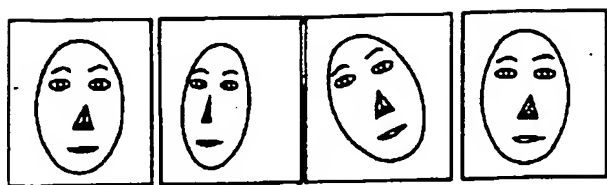
[Drawing 4]



[Drawing 5]



[Drawing 6]



基底画像 1 基底画像 2 基底画像 3 基底画像 4



線形和



☆ 特徴点

任意の方向の顔画像

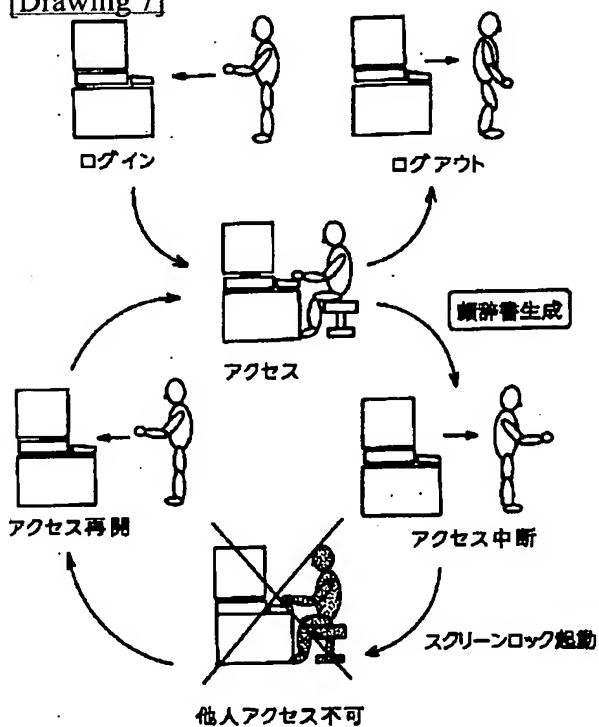
特徴点の座標 (x, y)

$$X = a1 \cdot x1 + a2 \cdot x2 + a3 \cdot x3 + a4 \cdot x4$$

$$Y = b1 \cdot y1 + b2 \cdot y2 + b3 \cdot y3 + b4 \cdot y4$$

4点以上で係数  $a1-a4$   $b1-b4$  を決定

[Drawing 7]



[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-232934

(43) 公開日 平成10年(1998) 9月2日

(51) Int.Cl.<sup>6</sup>

G 0 6 T 7/00

識別記号

F I

G 0 6 F 15/62

15/70

4 6 5 K

4 6 5 Z

審査請求 未請求 請求項の数6 O L (全 14 頁)

(21) 出願番号 特願平9-33704

(22) 出願日 平成9年(1997) 2月18日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 福井 和広

大阪府大阪市北区大淀中一丁目1番30号

梅田スカイビル タワーウエスト 株式会  
社東芝関西支社内

(72) 発明者 山口 修

大阪府大阪市北区大淀中一丁目1番30号

梅田スカイビル タワーウエスト 株式会  
社東芝関西支社内

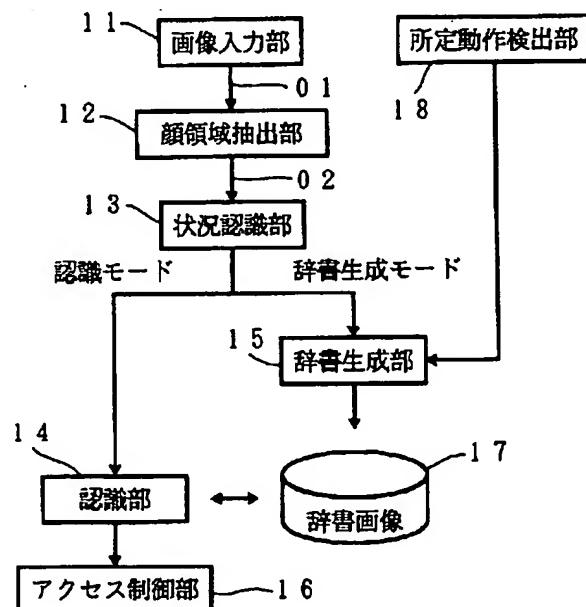
(74) 代理人 弁理士 薦田 璋子 (外1名)

(54) 【発明の名称】 顔画像登録装置及びその方法

(57) 【要約】

【課題】 使用者に負担かけず、かつ、不要な顔画像の登録を排除できて辞書精度を上げることができる顔画像登録装置を提供する。

【解決手段】 セキュリティ付き情報端末装置10は、画像入力部11、顔領域抽出部12、状況認識部13、認識部14、辞書生成部15、アクセス制御部16、辞書画像17、所定動作検出部18から構成され、所定動作検出部18において使用者の所定動作を検出した後、画像入力部11によって入力した画像から顔領域抽出部12が顔領域の画像を抽出して、辞書生成部15で辞書画像17に登録する価値があるか否かを判断する。



**【特許請求の範囲】**

【請求項1】人物の顔画像を登録している辞書手段を有する顔画像登録装置において、

前記人物の画像を入力する画像入力手段と、

前記画像入力手段によって入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出手段と、

前記顔領域抽出手段によって抽出された入力顔画像が、前記辞書手段に登録可能な顔画像であるか否かを判断する判断手段と、

前記判断手段によって登録可能と判断された入力顔画像を前記顔画像として前記辞書手段に登録させる登録手段からなることを特徴とする顔画像登録装置。

【請求項2】前記判断手段は、

前記顔領域抽出手段によって抽出された入力顔画像から顔の特徴量を抽出し、

この抽出された特徴量と、予め設定された基準とを比較して、

前記特徴量が前記基準以上であれば登録可能と判断することを特徴とする請求項1記載の顔画像登録装置。

【請求項3】前記人物の所定の動作を検出する動作検出手段を有し、

前記判断手段は、

前記動作検出手段が前記所定の動作を検出している間、または、その検出後一定期間の間に、前記顔領域抽出手段によって抽出された入力顔画像が前記辞書手段に登録するための前記顔画像として登録可能か否かを判断することを特徴とする請求項1記載の顔画像登録装置。

【請求項4】人物の顔画像を登録するための顔画像登録方法において、

前記人物の画像を入力する画像入力ステップと、

前記画像入力ステップにおいて入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出ステップと、

前記顔領域抽出ステップにおいて抽出された入力顔画像が、前記登録可能な顔画像か否かを判断する判断ステップと、

前記判断ステップにおいて登録可能と判断された入力顔画像を前記顔画像として登録させる登録ステップからなることを特徴とする顔画像登録方法。

【請求項5】前記判断ステップは、

前記顔領域抽出ステップにおいて抽出された入力顔画像から顔の特徴量を抽出し、

この抽出された特徴量と、予め設定された基準とを比較して、

前記特徴量が前記基準以上であれば登録可能と判断することを特徴とする請求項4記載の顔画像登録方法。

【請求項6】前記人物の所定の動作を検出する動作検出ステップを有し、

前記判断ステップは、

前記動作検出ステップにおいて前記所定の動作を検出している間、または、その検出後一定期間の間に、前記顔領域抽出ステップにおいて抽出された入力顔画像が前記顔画像として登録可能か否かを判断することを特徴とする請求項4記載の顔画像登録方法。

**【発明の詳細な説明】****【0001】**

【発明の属する技術分野】本発明は、入力された画像中から顔領域を抽出して自動的に顔を辞書登録する顔画像登録装置またはその方法に関する。

**【0002】**

【従来の技術】顔画像認識技術は、ヒューマンインタフェースあるいはセキュリティシステムの構築には不可欠な要素技術である。顔画像認識は、顔領域抽出、顔特徴点抽出、識別からなる。最近の技術動向に関しては、文献[塩野充、真田英彦“個人認証技術の最近の研究動向”，信学技報 OSF92-17.]が詳しい。従来の顔画像識別法は、大きく分ける以下の2つに分類できる。

【0003】1つは、目、鼻、口などの特徴点の位置、形状、サイズをパラメータ化して特徴ベクトルを生成し、予め登録されている対象人物の特徴ベクトルとの類似度を計算する方法である。最も類似度が高い辞書ベクトルを表す人物を当人と識別する。これらは構造解析的な手法に分類される。

【0004】もう1つの方法は、瞳、鼻などの特徴点を基準とした2次元affine変換などの幾何学変換により位置、サイズを正規化された画像と予め登録されている辞書正規化画像とのパターンの類似度に基づいた方法である。前者と同様に最も類似度が高い辞書画像を表す人物を当人と識別する。これらは従来の文字認識で実用化されている方法でパターン的な手法に分類できる。

【0005】上記の2つの手法のどちらの場合も、類似度としては画像間の相関値や特徴空間中でのユークリッド距離などの統計的な距離を用いる。この識別処理には文字認識で実用化されている様々なパターン認識理論、例えば部分空間法[エルッキ・オヤ著 小川秀光、佐藤誠訳、“パターン認識と部分空間法”産業図書(1986)]や複合類似度など方法が適用できる。識別法の構成は[舟久保登“パターン認識”共立出版(1991)][飯島泰蔵“パターン認識理論”森北出版(1989)]などに詳しい。

【0006】上記の顔認識の実際の適用に際しては、上記の顔認識処理に加えて使用者の登録、つまり登録者の顔辞書生成が不可欠になってくる。例えば、部分空間法を用いた手書き漢字認識では、各文字毎に数百枚の学習サンプル画像から辞書を生成していた。本発明で対象としている顔のような3次元物体は、文字に比べて照明条件、顔向き、表情変化などの形状や輝度の変動が大きく、さらに大量の学習サンプル画像が必要となってくる[村瀬洋、シェリー・ナイヤー，“2次元照合による3

次元物体認識”，信学論(D-II) J77-D-II, 11, pp. 2179-2187, 1994.]。

【0007】従来の辞書生成は、文献〔赤松茂，佐々木努，深町映夫，末永康仁，“濃淡画像マッチングによるロバストな正面顔の識別法”，信学論(D-II)，J76-D-I, 7, pp. 1363-1373, 1993.〕、〔小松良江，有木康雄，“部分空間法を用いた向きによらない顔の切り出し認識”，PRU95-191, pp. 7-14, 1996.〕、〔M. Turk, A. P. Pentland: "Face recognition using eigenfaces", Proc. CVPR 11, pp. 453-458, 1993.〕、〔Alex Pentland, Baback Moghaddam, Thad Starner, "View-based and modular eigenspaces for face recognition", CVPR '94, pp. 84-91, 1994.〕などに見られるように、良好な照明条件下で所定、例えば、正面、左右15°、上下15°刻の方向に登録者の顔を向けて撮影した画像から人手により顔の切り出し及び画像の選択を行っていた。

【0008】顔はその日の体調や表情変化、髪、髭の影響などの生理的な変化が生じ易い。したがって高い認識率を実現するためには、顔辞書は一定期間毎に更新する必要がある。しかし頻繁な辞書更新は使用者の負担を高めることになる。そこで顔認識を使い易くするためには、使用者に如何に意識させないで負担を軽減して登録するかが重要になってくる。

【0009】例えば、情報端末における作業を一時中断して立ち去る場合に、他人のアクセスを防止する機能を考える。予め生成した辞書を用いることも考えられるが、本人同定、他人排除率を上げるためには、できるだけ直前に生成された辞書画像を用いる方が有効である。しかし離席の度に辞書登録することは、使用者の負担を高くするため、使用者が情報機器をアクセス中に意識させないで辞書登録する。

【0010】ところが登録者に意識させないで収集した学習サンプルは、通常は目を閉じた場合や、表情が大きく変化している場合などの学習データとして有効でないデータを含んでいる。これらの誤り画像が含まれている学習サンプルから辞書生成を行うと辞書の精度が低下してしまう。辞書精度を上げるためには、学習データ数を増やす必要があり、人手による選択ではリアルタイム辞書生成に対処できない。

【0011】

【発明が解決しようとする課題】従来、顔辞書の生成は、事前に良好な照明条件下で所定、例えば、正面、左右15°、上下15°毎の方向に登録者の顔を向けさせて撮影した顔画像を用いて人手により行っていた。顔は時間と共に変化する。したがって高い認識率を実現するためには顔辞書は一定期間毎に更新する必要がある。しかし頻繁な辞書更新を行うことは、使用者の負担を高めることになる。

【0012】辞書生成のための自動学習サンプル収集に

関しては、既に出願済みの方法〔特願平8-61463号〕を適用すれば実現できる。しかし、依然として被登録者に意識させないで収集した学習データには辞書生成に有効でないデータが多数含まれるという課題が残る。辞書精度を上げるためには、学習データ数を増やす必要があり、人手による選択ではリアルタイム辞書生成が実現困難である。

【0013】そこで、本発明は、使用者に負担かけず、かつ、不要な顔画像の登録を排除できて辞書精度を上げることができる顔画像登録装置及びその方法を提供する。

【0014】

【課題を解決するための手段】本発明は、人物の顔画像を登録している辞書手段を有する顔画像登録装置において、前記人物の画像を入力する画像入力手段と、前記画像入力手段によって入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出手段と、前記顔領域抽出手段によって抽出された入力顔画像が、前記辞書手段に登録可能な顔画像であるか否かを判断する判断手段と、前記判断手段によって登録可能と判断された入力顔画像を前記顔画像として前記辞書手段に登録させる登録手段からなることを特徴とする顔画像登録装置である。

【0015】また、本発明は、人物の顔画像を登録するための顔画像登録方法において、前記人物の画像を入力する画像入力ステップと、前記画像入力ステップにおいて入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出ステップと、前記顔領域抽出ステップにおいて抽出された入力顔画像が、前記登録可能な顔画像か否かを判断する判断ステップと、前記判断ステップにおいて登録可能と判断された入力顔画像を前記顔画像として登録させる登録ステップからなることを特徴とする顔画像登録方法である。

【0016】上記発明であると、例えば、抽出された顔領域から求まる特徴量を予め設定された基準とを比較することにより学習データとして有効かの判断を行う。特徴量としては、瞳、鼻穴、口端などの顔特徴点の大きさ、位置情報から求まる量あるいはこれらの点を基準にして正規化を施した画像を用いる。これにより大きく表情が変化した画像、目を閉じた画像、口を開けた画像などを学習画像から自動的に取り除くことが可能になる。

【0017】さらにマウス操作、キーボード操作、ボタン操作などの所定の操作を検出して、操作中に入力された画像のみを対象として辞書生成を行う。操作中のみを対象とすることで無駄な学習データ収集の時間が減り効率的である。また横を向いている場合などの辞書として有効でない画像を収集する可能性も減る。本発明によれば、入力された画像中から顔領域を抽出して自動的に辞書登録する装置及び方法を実現できる。

【0018】

【発明の実施の形態】以下、本発明の一実施例を図1から図7に基づいて説明する。

【0019】パソコン等の情報端末装置において、作業中を一時中断して端末を離れる際に他人のアクセスを防止することは必要な機能である。そのため、本実施例では、離席時に他人のアクセスを防止する機能を搭載した情報機器端末装置10を例にとって説明する。

【0020】図1は、本実施例に係るセキュリティ付き情報端末装置10の概略を示すブロック図である。

【0021】本装置10は、画像入力部11、顔領域抽出部12、状況認識部13、認識部14、辞書生成部15、アクセス制御部16、辞書画像17、所定動作検出部18から構成されている。

【0022】(画像入力部11) 画像入力部11は、認識対象となる人物の画像を入力するためのものであり、例えばTVカメラからなる。この画像入力部11から入力された画像01はA/D変換器によりデジタル化されて顔領域抽出部12に送られる。例えば、TVカメラは情報端末装置10のモニタの下部に設置される。あるいはモニタの四角に設置しても良い。

【0023】(顔領域抽出部12) 顔領域抽出部12は、画像入力部11から送られてきた入力画像から顔領域画像02を常時抽出し続ける。本実施例では、予め登録された標準顔画像(テンプレート)を全画面に渡って移動させながら相関値を計算し最も高い相関値をもっている領域を顔領域とする。相関値の局所最大点を顔領域候補とする。この顔領域候補における相関値が設定された閾値より低い場合は、顔が存在しないとする。具体的には閾値として最大相関値を100として30に設定する。顔の向き変化に対応するために複合類似度などにより複数のテンプレートを用いるとさらに安定に顔領域を抽出できる。この処理は先に述べたカラー情報に基づく抽出法に置き換えて良い。

【0024】(状況認識部13) 状況認識部13では、顔領域抽出結果の時間的な変化から使用者の状態(作業中、離席、着席)を識別して辞書生成モードと認識モードの切替えを行う。

【0025】図2に示すように顔領域が検出され始めた時点T0からT1までの期間に認識モードに設定する。このモードでは後で述べる認識処理により本人同定処理を行う。T1時点で本人と同定された場合はアクセス可能になる。

【0026】T1時点以降は辞書生成モードに切り替わり所定動作検出部18からの辞書生成信号が入力された時に学習データの収集が行われる。一定期間顔領域を検出できなくなったら離席したと判断して即座にスクリーンロックを起動する。あるいは使用者に光、音声などでその主旨を伝えて使用者の反応を見てから、スクリーンロックを起動しても良い。ここで時刻T1は対象機種やアクセス動作の種類により変更可能である。

【0027】(認識部14) 認識部14は、顔特徴点抽出部14a、正規化画像生成部14b、パターンマッチング部14cからなる。図3に認識部14のブロック図を示す。

【0028】顔特徴点抽出部14aでは、抽出された顔領域内から瞳、鼻、口橋などの特徴点を抽出する。既に出願している形状情報とパターン情報を組み合わせた方法[特願平8-61463号]が適用可能である。

【0029】この方法の基本的な考えは、位置精度の高い形状情報により特徴点の候補を求め、それをパターン照合で検証するというものである。本方法は形状情報により位置決めを行うので高い位置精度を期待できる。また候補群からの正しい特徴点の選択に、マルチテンプレートをを用いたマッチングを適用しているために特徴点の形状輝度の変動に対してロバストである。処理速度に関しては、計算コストの少ない分離度フィルターで絞り込んだ候補に対してのみパターン照合するので全体をパターン照合する方法に比べ計算量の大幅な削減が実現できる。

【0030】この他にも、エッジ情報に基づく方法[A. L. Yuille, " Feature extraction from faces using deformable templates ", IJCV, vol. 8: 2, pp. 99-111, 1992.] [坂本静生, 宮尾陽子, 田島譲二, "顔画像からの目の特徴点抽出", 信学論 D-11, Vol. J76-D-1 I, No. 8, pp. 1796-1804, August, 1993.] や固有空間法を適用したEigen feature 法[Alex Pentland, Baba ck Moghaddam, Thad Starner, " View-based and modular eigenspaces for face recognition ", CVPR '94, p p. 84-91, 1994.]、カラー情報に基づく方法[佐々木努、赤松茂、末永康仁, "顔画像認識のための色情報を用いた顔の位置合わせ方", IE91-2, pp. 9-15, 1991.] が適用可能である。

【0031】正規化画像生成部14bでは、特徴点を基準にして正規化を施す。図4に瞳、鼻穴を基準にした正規化処理の例を示す。ベクトルE1E2の向きを平行に補正してさらにc1, c2の上から1/3の点CPを正規化画像の中心座標、横幅はベクトルE1E2の2倍、縦幅はベクトルc1c2の2倍に設定する。

【0032】パターンマッチング部14cでは、正規化画像と辞書画像に蓄えられている顔画像と比較してパターン類似度を求める。パターン類似度が基準値より高い場合には、本人であると同定する。基準より小さい場合には、他人とする。パターンマッチングの方法としては、相関法、部分空間法、複合類似度法などが適用可能である。

【0033】(所定動作検出部18) 所定動作検出部18は、使用者がキーボード、ボタンプッシュ、または、マウスの操作などの所定の動作を行っているか否かを検出する。検出結果に基づいて辞書生成部15を制御する。

【0034】具体的には、使用者がキーボード操作、ポタンブッシュ、マウスの操作などの所定の動作を行っていることを検知すると、画像入力手段11に使用者の顔の正面画像が写っている可能性が高いため、辞書登録するのに好適であるために所定の動作を行っている間は辞書生成信号を辞書生成部16に送信する。

【0035】(辞書生成部16) 辞書生成部16は、顔特徴点抽出部14a、フレーム評価部16a、正規化画像生成部14b、学習画像記憶部16b及び主成分分析部16cからなる。ここで顔特徴点抽出部14aと正規化画像生成部14bは、認識部14と共通で使用する。図5に辞書生成部16のブロック図を示す。

【0036】辞書生成は、所定動作検出部18から辞書生成信号が入力されると学習データの収集が行われる。

【0037】フレーム評価部16bでは、抽出された顔特徴点の情報から抽出された顔領域画像02が辞書生成に有効か否かを判断する。ここで3つの評価基準を適用するが、個別あるいは組合せで適用してもよい。

【0038】(i) 予め各特徴点の位置関係を規定しておき、この位置関係が崩れた場合には、学習データとして有効でないと判断する。例えば、両瞳、鼻穴、口端の位置関係などである。

【0039】(ii) 辞書画像17に蓄えられた固有顔画像とパターン類似度を評価にする。類似度が基準より小さい場合には有効でないと判断する。基準を高くする程、学習データの均一性を高めることになる。しかし高過ぎると顔向きなどの変化に対応できなくなる。

【0040】あるいは辞書画像17に蓄えられた固有顔画像の代わりに、事前に正しい特徴点で正規化された学習画像から生成した辞書画像を用いても良い。

【0041】さらに顔向きまで考慮すると極端に横を向いた顔などを除くことが可能である。この場合には、予め方向1～Nの顔向き毎に準備してある顔辞書パターン

$$X = a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot x_4 \quad (1)$$

$$Y = b_1 \cdot y_1 + b_2 \cdot y_2 + b_3 \cdot y_3 + b_4 \cdot y_4 \quad (2)$$

$$Res = \sqrt{(X - x')^2 + (Y - y')^2} \quad (3)$$

誤差Resは新しい特徴点の座標誤差を示している。誤差が閾値より大きい場合は、抽出位置に誤りが生じているので、このデータは学習データとして有効でない。

【0047】上記(i)～(iii)の評価基準以外にも顔領域の平均輝度や輝度の分散が基準範囲に入っているか否かで判断することも可能である。

【0048】学習画像記憶部16bは、上記有効であると判断された学習データを蓄える。蓄えられた学習データ数が規定の数に達したら主成分分析部16cに学習画像を送る。

【0049】主成分分析部16cでは、学習画像記憶部16bに蓄えられた画像データに対して主成分分析(K

との比較を行って、所定の顔向きの辞書に一致する場合に辞書生成に有効であると判断する。

【0042】(iii) 各フレーム毎に抽出された特徴点の関係を調べて目を閉じた学習データを検出する。新しく抽出された特徴点の座標は、既にその前のフレーム1～4で抽出された特徴点の座標の線形結合で表現できる。したがって全てのフレームで正しく特徴点が抽出されていれば、線形結合した際の誤差が小さくなる。逆に目を閉じて眉と瞳を間違えた場合などには、誤差が大きくなる。

【0043】これについて詳細に図6を用いて説明する。

【0044】正射影モデルを仮定すると、3次元モデルを持たずに、任意の方向から見た顔特徴点の2次元座標を、式(1)(2)に示すように4枚の画像上で対応付けられた特徴点の2次元座標の線形結合により表現できる[S. Ullman, R. Basri: "Recognition by Linear Combinations of Models, IEEE Trans. PAMI, Vol. 13, No. 10, pp. 992-1006, 1991.][向川康博、中村裕一、大田友一、"2枚の顔写真を用いた任意方向の顔画像の生成"、情処論、Vol. 37, No. 4, pp. 635-644, 1996.]。この性質を利用して動画像列において既に抽出されている4枚のフレームにおける正しい特徴点の2次元座標値( $x_1, y_1$ )～( $x_4, y_4$ )から新しく検出された特徴点の2次元座標値( $X, Y$ )を検証する。

【0045】4フレームの画像に対する特徴点の2次元座標と新しい座標( $X, Y$ )から線形結合係数 $a_i, b_i, (i=1, 4)$ を最小自乗法により求める。今度は逆に得られた結合係数から近似座標を( $x', y'$ )を計算して( $X, Y$ )との誤差Resを計算する。

【0046】

【数1】

L展開)適用して固有画像を求める。固有値が大きい方から上位N個の固有画像を辞書画像17に蓄える。

【0050】上記実施例では、動画像からリアルタイムで辞書生成を行う例を説明したが、動画像から所定動作検出部からの検出信号が来たフレームだけを一旦メモリに蓄えてから行うことも可能である。この場合は、辞書生成に有効である可能性が高いフレームだけを対象にするために、処理の高速化と安定性が計れる。

【0051】また、上記実施例では、識別にパター的な手法である部分空間法を適用しているが、構造的な手法に置き換えても良い。

【0052】(本装置10の動作の説明)次に、本実施

例の動作を図7に沿って説明する。

【0053】最初のアクセスは、パスワードなどを入力する通常のログインを行う。あるいは、事前に登録した使用者の辞書を用いて自動ログインを行うことも可能である。

【0054】まず画像入力部11から入力された画像から顔領域抽出部12において顔領域が抽出されて状況認識部13に送られる。また所定動作検出部18では、マウス操作、キーボード操作、ボタン操作などの所定動作検出信号を状況認識部に送る。状況認識部13では、顔領域の時間的な有無に基づいて認識モードと辞書生成モードの切替えを行う。また所定動作検出信号に基づいて学習データを収集するか否かを制御する。

【0055】まず時刻T0からT1までは、認識モードに設定され、前回の使用者本人であるかを同定する。

【0056】(i) 認識モード

顔領域は、顔特徴点抽出部14aに送られて瞳、鼻穴、口端が検出される。次にこれらの特徴点を基準に正規化が施される。正規化画像は、パターンマッチング部14cで前回使用者の辞書画像との類似度が計算され、類似度が基準値より高い場合に前回使用していた本人と同定されてアクセス可能になる。基準値より小さい場合はアクセスを禁止する。

【0057】次に時刻T1からは、辞書生成モードに設定される。

【0058】(ii) 辞書生成モード

所定動作検出部18が所定動作を検出して所定動作検出信号がONになっている場合に学習データを収集して辞書画像を生成する。連続して入力される画像から抽出された顔領域画像は顔特徴点抽出部14aに送られて瞳、鼻穴、口端が検出される。抽出された特徴点情報はフレーム評価部16aに送られて、例えば式(3)から誤差が計算される。誤差が閾値より小さい場合には学習データとして有効であると判断され学習画像記憶部16bに蓄えられる。学習画像記憶部16bに規定枚数以上の学習データが蓄えられたら学習データは主成分分析部16cに送られて辞書画像が生成される。

【0059】使用者が離席すると即座にスクリーンロックを起動する。再び使用者が着席すると最初の処理に戻る。

【0060】なお、上記説明では所定動作検出部18が所定動作を検出して所定動作検出信号がONになっている場合に辞書画像を生成していたが、さらに、所定動作検出信号がOFFにあった後、一定期間(例えば、1分間)の間は続けて辞書画像を生成してもよい。

【0061】(変更例)次に、情報端末のセキュリティだけではなく、情報端末のサービスにおける実施例について説明する。

【0062】例えば、顔を登録する場合に、顔領域から得られる特徴量だけではなく、情報端末の操作履歴も同

時に記憶することを考える。顔の情報と同時に登録しておくことによって、その人間の識別と同時に、端末を操作していたときに、どのような情報を求めていたのかを履歴から抽出できる。

【0063】情報検索を例とすると、使用者は検索式等を入力する作業を行う。個人識別のために必要な情報に加え、その検索式、検索内容の履歴をデータベースに蓄えておく。蓄えられたデータをもとに、使用者の興味ある内容を推定、抽出し、新たな情報検索を行って保持しておく。

【0064】そして使用者が再び、情報端末へのログイン、操作復帰を行った場合、端末に新たに検索した所望と考えられる情報の提供を自動的に行うサービスが可能となる。

【0065】これらの履歴情報は顔の登録情報とリンクして蓄えられることにより、いつこのような情報を検索していたかというタイムスタンプの代わりになり、これに基づいて、いつの時点の顔の状態に近いのかを認識して、その当時の履歴情報に基づいた情報を提示しても良い。

【0066】また、顔の登録の際の検証機構として、表情検出、認識を用いれば、人間の感情状態に応じた情報の操作履歴をとることも可能であり、平常時の辞書、笑ったときの辞書、怒っているときの辞書などを生成することもできる。新たに顔認識を行う際に、表情情報に対応した情報提供を行っても良い。

【0067】これらは顔登録の更新と情報操作履歴の更新を継続的に同時に行うことにより、より確かな情報提供が可能になる。

【0068】これらの実施例は情報端末を例に説明したが、例えばTV、電話など家電製品に置き換えても良い。

【0069】TVの場合には、リモコンに超小型カメラを搭載する。所定の動作はリモコンのボタンを押している動作と定義する。電話の場合には、受話器に超小型CCDが搭載して受話器を上げる動作を所定の動作と定義する。

【0070】さらに、本実施例の装置10を現金自動預け払い機(以下、ATMという)に使用してもよい。

【0071】すなわち、使用者の本人確認のためにこの装置10を使用することにより、他人による取り忘れ、現金の横取りなどの不正使用を確実に防止できる。この場合には、画像入力部11を、ATMの操作パネルの近傍に配置して使用者の顔が確実に写るようにする。また、顔画像と同時に口座番号や更新登録時刻を登録しておいてもよい。

【0072】

【発明の効果】以上、本発明によれば、入力された画像中から新規顔画像を抽出して、これが辞書登録に値する顔画像か否かを判断して、有効と判断した入力顔画像の

みを登録するために、不要な顔画像の登録を排除できて辞書精度を上げることができる。

【図面の簡単な説明】

【図1】本実施例の装置のブロック図である。

【図2】本人同定モード及び辞書生成モードの時間的な変化を示す図である。

【図3】認識部のブロック図である。

【図4】瞳、鼻穴を基準にした正規化処理の例を示す図である。

【図5】辞書生成部のブロック図である。

【図6】線形結合した際の誤差を使用した説明図であ

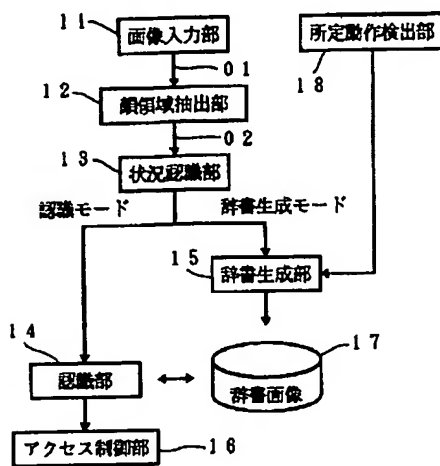
る。

【図7】本実施例の動作を説明する図である。

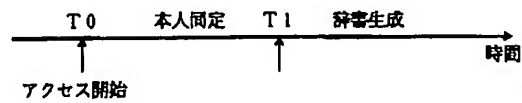
【符号の説明】

- 11 画像入力部
- 12 顔領域抽出部
- 13 状況認識部
- 14 認識部
- 15 辞書生成部
- 16 アクセス制御部
- 17 辞書画像
- 18 所定動作検出部

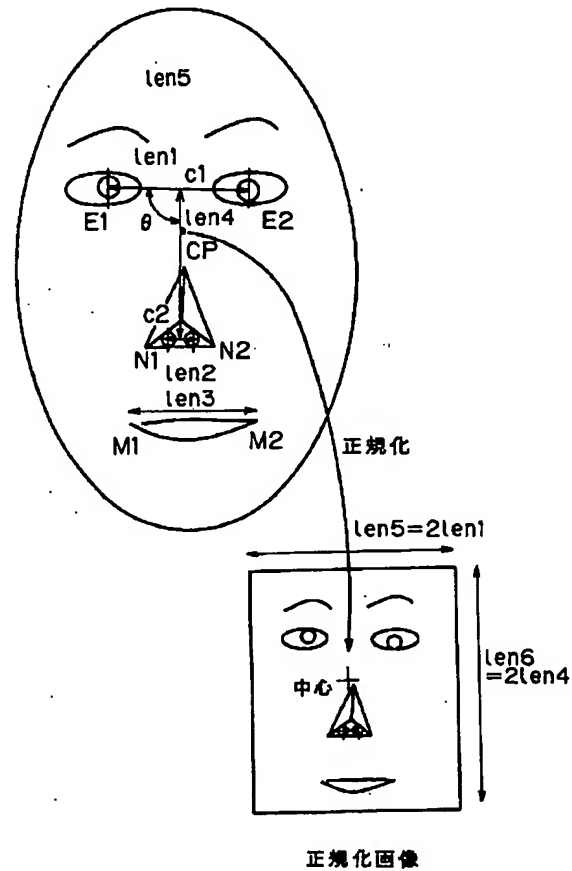
【図1】



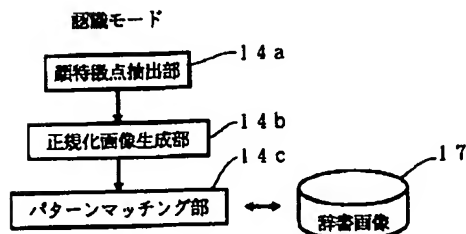
【図2】



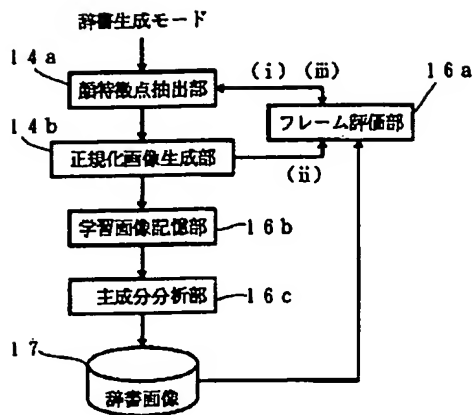
【図4】



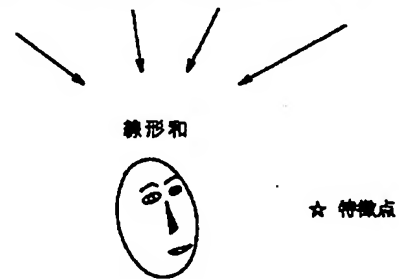
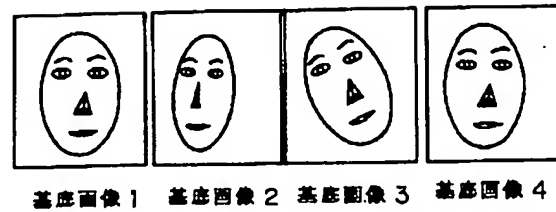
【図3】



【図5】



【図6】



任意の方向の顔画像

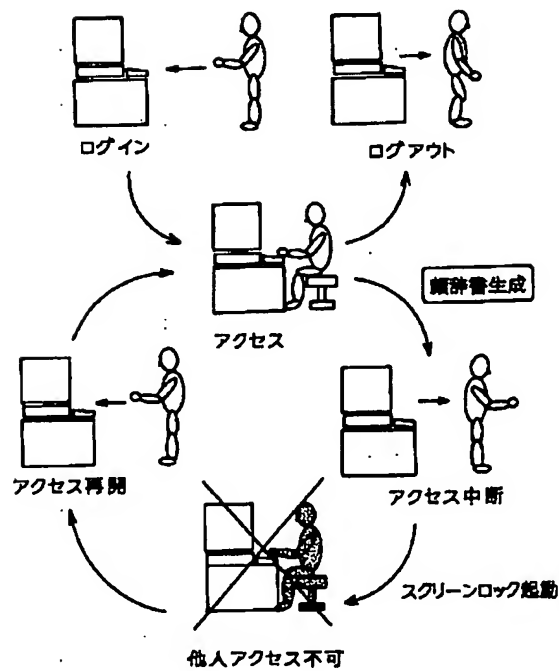
特徴点の座標 (x, y)

$$X = a1 \cdot x1 + a2 \cdot x2 + a3 \cdot x3 + a4 \cdot x4$$

$$Y = b1 \cdot y1 + b2 \cdot y2 + b3 \cdot y3 + b4 \cdot y4$$

4点以上で係数 a1-a4 b1-b4 を決定

【図7】



## 【手続補正書】

【提出日】平成9年2月25日

## 【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正内容】

【書類名】明細書

【発明の名称】顔画像登録装置及びその方法

【特許請求の範囲】

【請求項1】人物の顔画像を登録している辞書手段を有する顔画像登録装置において、前記人物の画像を入力する画像入力手段と、前記画像入力手段によって入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出手段と、前記顔領域抽出手段によって抽出された入力顔画像が、前記辞書手段に登録可能な顔画像であるか否かを判断する判断手段と、前記判断手段によって登録可能と判断された入力顔画像を前記顔画像として前記辞書手段に登録させる登録手段からなることを特徴とする顔画像登録装置。

【請求項2】前記判断手段は、前記顔領域抽出手段によって抽出された入力顔画像から顔の特徴量を抽出し、この抽出された特徴量と、予め設定された基準とを比較して、前記特徴量が前記基準以上であれば登録可能と判断することを特徴とする請求項1記載の顔画像登録装置。

【請求項3】前記人物の所定の動作を検出する動作検出手段を有し、前記判断手段は、前記動作検出手段が前記所定の動作を検出している間、または、その検出後一定期間の間に、前記顔領域抽出手段によって抽出された入力顔画像が前記辞書手段に登録するための前記顔画像として登録可能か否かを判断することを特徴とする請求項1記載の顔画像登録装置。

【請求項4】人物の顔画像を登録するための顔画像登録方法において、前記人物の画像を入力する画像入力ステップと、前記画像入力ステップにおいて入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出ステップと、前記顔領域抽出ステップにおいて抽出された入力顔画像が、前記登録可能な顔画像か否かを判断する判断ステップと、前記判断ステップにおいて登録可能と判断された入力顔画像を前記顔画像として登録させる登録ステップからなることを特徴とする顔画像登録方法。

【請求項5】前記判断ステップは、

前記顔領域抽出ステップにおいて抽出された入力顔画像から顔の特徴量を抽出し、

この抽出された特徴量と、予め設定された基準とを比較して、

前記特徴量が前記基準以上であれば登録可能と判断することを特徴とする請求項4記載の顔画像登録方法。

【請求項6】前記人物の所定の動作を検出する動作検出ステップを有し、

前記判断ステップは、

前記動作検出ステップにおいて前記所定の動作を検出している間、または、その検出後一定期間の間に、前記顔領域抽出ステップにおいて抽出された入力顔画像が前記顔画像として登録可能か否かを判断することを特徴とする請求項4記載の顔画像登録方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、入力された画像中から顔領域を抽出して自動的に顔を辞書登録する顔画像登録装置またはその方法に関する。

【0002】

【従来の技術】顔画像認識技術は、ヒューマンインタフェースあるいはセキュリティシステムの構築には不可欠な要素技術である。顔画像認識は、顔領域抽出、顔特徴点抽出、識別からなる。最近の技術動向に関しては、文献[塩野充、真田英彦“個人認証技術の最近の研究動向”，信学技報 OSF92-17.]が詳しい。従来の顔画像識別法は、大きく分ける以下の2つに分類できる。

【0003】1つは、目、鼻、口などの特徴点の位置、形状、サイズをパラメータ化して特徴ベクトルを生成し、予め登録されている対象人物の特徴ベクトルとの類似度を計算する方法である。最も類似度が高い辞書ベクトルを表す人物を当人と識別する。これらは構造解析的な手法に分類される。

【0004】もう1つの方法は、瞳、鼻などの特徴点を基準とした2次元affine変換などの幾何学変換により位置、サイズを正規化された画像と予め登録されている辞書正規化画像とのパターンの類似度に基づいた方法である。前者と同様に最も類似度が高い辞書画像を表す人物を当人と識別する。これらは従来の文字認識で実用化されている方法でパターン的な手法に分類できる。

【0005】上記の2つの手法のどちらの場合も、類似度としては画像間の相関値や特徴空間中でのユークリッド距離などの統計的な距離を用いる。この識別処理には文字認識で実用化されている様々なパターン認識理論、例えば部分空間法[エルッキ・オヤ著 小川秀光、佐藤誠訳、“パターン認識と部分空間法”産業図書(1986)]や複合類似度など方法が適用できる。識別法の構成は[舟久保登“パターン認識”共立出版(1991)] [飯島泰蔵“パターン認識理論”森北出版(1989)]などに

詳しい。

【0006】上記の顔認識の実際の適用に際しては、上記の顔認識処理に加えて使用者の登録、つまり登録者の顔辞書生成が不可欠になってくる。例えば、部分空間法を用いた手書き漢字認識では、各文字毎に数百枚の学習サンプル画像から辞書を生成していた。本発明で対象としている顔のような3次元物体は、文字に比べて照明条件、顔向き、表情変化などの形状や輝度の変動が大きく、さらに大量の学習サンプル画像が必要となってくる〔村瀬洋，シェリー・ナイヤー，“2次元照合による3次元物体認識”，信学論（D-II）J77-D-II，11，pp. 2179-2187，1994.〕。

【0007】従来の辞書生成は、文献〔赤松茂，佐々木努，深町映夫，末永康仁，“濃淡画像マッチングによるロバストな正面顔の識別法”，信学論（D-II），J76-D-I，7，pp. 1363-1373，1993.〕、〔小松良江，有木康雄，“部分空間法を用いた向きによらない顔の切り出し認識”，PRU95-191，pp. 7-14，1996.〕、〔M. Turk，A. P. Pentland：“Face recognition using eigenfaces”，Proc. CVPR 11，pp. 453-458，1993.〕、〔Alex Pentland，Baback Moghaddam，Thad Starner，“View-based and modular eigenspaces for face recognition”，CVPR '94，pp. 84-91，1994.〕などに見られるように、良好な照明条件下で所定、例えば、正面、左右15°、上下15°刻の方向に登録者の顔を向けて撮影した画像から人手により顔の切り出し及び画像の選択を行っていた。

【0008】顔はその日の体調や表情変化、髪、髭の影響などの生理的な変化が生じ易い。したがって高い認識率を実現するためには、顔辞書は一定期間毎に更新する必要がある。しかし頻繁な辞書更新は使用者の負担を高めることになる。そこで顔認識を使い易くするためには、使用者に如何に意識させないで負担を軽減して登録するかが重要になってくる。

【0009】例えば、情報端末における作業を一時中断して立ち去る場合に、他人のアクセスを防止する機能を考える。予め生成した辞書を用いることも考えられるが、本人同定、他人排除率を上げるためには、できるだけ直前に生成された辞書画像を用いる方が有効である。しかし離席の度に辞書登録することは、使用者の負担を高くするため、使用者が情報機器をアクセス中に意識させないで辞書登録する。

【0010】ところが登録者に意識させないで収集した学習サンプルは、通常は目を閉じた場合や、表情が大きく変化している場合などの学習データとして有効でないデータを含んでいる。これらの誤り画像が含まれている学習サンプルから辞書生成を行うと辞書の精度が低下してしまう。辞書精度を上げるためには、学習データ数を増やす必要があり、人手による選択ではリアルタイム辞書生成に対処できない。

【0011】

【発明が解決しようとする課題】従来、顔辞書の生成は、事前に良好な照明条件下で所定、例えば、正面、左右15°、上下15°毎の方向に登録者の顔を向けさせて撮影した顔画像を用いて人手により行っていた。顔は時間と共に変化する。したがって高い認識率を実現するためには顔辞書は一定期間毎に更新する必要がある。しかし頻繁な辞書更新を行うことは、使用者の負担を高めることになる。

【0012】辞書生成のための自動学習サンプル収集に関しては、既に出願済みの方法〔特願平8-61463号〕を適用すれば実現できる。しかし、依然として被登録者に意識させないで収集した学習データには辞書生成に有効でないデータが多数含まれるという課題が残る。辞書精度を上げるためには、学習データ数を増やす必要があり、人手による選択ではリアルタイム辞書生成が実現困難である。

【0013】そこで、本発明は、使用者に負担かけず、かつ、不要な顔画像の登録を排除できて辞書精度を上げることができる顔画像登録装置及びその方法を提供する。

【0014】

【課題を解決するための手段】本発明は、人物の顔画像を登録している辞書手段を有する顔画像登録装置において、前記人物の画像を入力する画像入力手段と、前記画像入力手段によって入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出手段と、前記顔領域抽出手段によって抽出された入力顔画像が、前記辞書手段に登録可能な顔画像であるか否かを判断する判断手段と、前記判断手段によって登録可能と判断された入力顔画像を前記顔画像として前記辞書手段に登録させる登録手段からなることを特徴とする顔画像登録装置である。

【0015】また、本発明は、人物の顔画像を登録するための顔画像登録方法において、前記人物の画像を入力する画像入力ステップと、前記画像入力ステップにおいて入力された画像中から前記人物の顔領域の画像である入力顔画像を抽出する顔領域抽出ステップと、前記顔領域抽出ステップにおいて抽出された入力顔画像が、前記登録可能な顔画像か否かを判断する判断ステップと、前記判断ステップにおいて登録可能と判断された入力顔画像を前記顔画像として登録させる登録ステップからなることを特徴とする顔画像登録方法である。

【0016】上記発明であると、例えば、抽出された顔領域から求まる特徴量を予め設定された基準とを比較することにより学習データとして有効かの判断を行う。特徴量としては、瞳、鼻穴、口端などの顔特徴点の大きさ、位置情報から求まる量あるいはこれらの点を基準にして正規化を施した画像を用いる。これにより大きく表情が変化した画像、目を閉じた画像、口を開けた画像な

どを学習画像から自動的に取り除くことが可能になる。

【0017】さらにマウス操作、キーボード操作、ボタン操作などの所定の操作を検出して、操作中に入力された画像のみを対象として辞書生成を行う。操作中のみを対象とすることで無駄な学習データ収集の時間が減り効率的である。また横を向いている場合などの辞書として有効でない画像を収集する可能性も減る。本発明によれば、入力された画像中から顔領域を抽出して自動的に辞書登録する装置及び方法を実現できる。

【0018】

【発明の実施の形態】以下、本発明の一実施例を図1から図7に基づいて説明する。

【0019】パソコン等の情報端末装置において、作業中を一時中断して端末を離れる際に他人のアクセスを防止することは必要な機能である。そのため、本実施例では、離席時に他人のアクセスを防止する機能を搭載した情報機器端末装置10を例にとって説明する。

【0020】図1は、本実施例に係るセキュリティ付き情報端末装置10の概略を示すブロック図である。

【0021】本装置10は、画像入力部11、顔領域抽出部12、状況認識部13、認識部14、辞書生成部15、アクセス制御部16、辞書画像17、所定動作検出部18から構成されている。

【0022】(画像入力部11) 画像入力部11は、認識対象となる人物の画像を入力するためのものであり、例えばTVカメラからなる。この画像入力部11から入力された画像01はA/D変換器によりデジタル化されて顔領域抽出部12に送られる。例えば、TVカメラは情報端末装置10のモニタの下部に設置される。あるいはモニタの四角に設置しても良い。

【0023】(顔領域抽出部12) 顔領域抽出部12は、画像入力部11から送られてきた入力画像から顔領域画像02を常時抽出し続ける。本実施例では、予め登録された標準顔画像(テンプレート)を全画面に渡って移動させながら相関値を計算し最も高い相関値をもっている領域を顔領域とする。相関値の局所最大点を顔領域候補とする。この顔領域候補における相関値が設定された閾値より低い場合は、顔が存在しないとする。具体的には閾値として最大相関値を100として30に設定する。顔の向き変化に対応するために複合類似度などにより複数のテンプレートを用いるとさらに安定に顔領域を抽出できる。この処理は先に述べたカラー情報に基づく抽出法に置き換えて良い。

【0024】(状況認識部13) 状況認識部13では、顔領域抽出結果の時間的な変化から使用者の状態(作業中、離席、着席)を識別して辞書生成モードと認識モードの切替えを行う。

【0025】図2に示すように顔領域が検出され始めた時点T0からT1までの期間に認識モードに設定する。このモードでは後で述べる認識処理により本人同定処理

を行う。T1時点で本人と同定された場合はアクセス可能になる。

【0026】T1時点以降は辞書生成モードに切替わり所定動作検出部18からの辞書生成信号が入力された時に学習データの収集が行われる。一定期間顔領域を検出できなくなったら離席したと判断して即座にスクリーンロックを起動する。あるいは使用者に光、音声などでその主旨を伝えて使用者の反応を見てから、スクリーンロックを起動しても良い。ここで時刻T1は対象機種やアクセス動作の種類により変更可能である。

【0027】(認識部14) 認識部14は、顔特徴点抽出部14a、正規化画像生成部14b、パターンマッチング部14cからなる。図3に認識部14のブロック図を示す。

【0028】顔特徴点抽出部14aでは、抽出された顔領域内から瞳、鼻、口橋などの特徴点を抽出する。既に出願している形状情報とパターン情報を組み合わせた方法[特願平8-61463号]が適用可能である。

【0029】この方法の基本的な考えは、位置精度の高い形状情報により特徴点の候補を求め、それをパターン照合で検証するというものである。本方法は形状情報により位置決めを行うので高い位置精度を期待できる。また候補群からの正しい特徴点の選択に、マルチテンプレートを用いたマッチングを適用しているために特徴点の形状輝度の変動に対してロバストである。処理速度に関しては、計算コストの少ない分離度フィルターで絞り込んだ候補に対してのみパターン照合するので全体をパターン照合する方法に比べ計算量の大幅な削減が実現できる。

【0030】この他にも、エッジ情報に基づく方法[A. L. Yuille, "Feature extraction from faces using deformable templates", IJCV, vol. 8: 2, pp. 99-111, 1992.] [坂本静生, 宮尾陽子, 田島譲二, "顔画像からの目の特徴点抽出", 信学論 D-II, Vol. J76-D-I I, No. 8, pp. 1796-1804, August, 1993.] や固有空間法を適用したEigen feature 法[Alex Pentland, Baba ck Moghaddam, Thad Starner, "View-based and modular eigenspaces for face recognition", CVPR '94, p p. 84-91, 1994.], カラー情報に基づく方法[佐々木努, 赤松茂, 末永康仁, "顔画像認識のための色情報を用いた顔の位置合わせ方", IE91-2, pp. 9-15, 1991.] が適用可能である。

【0031】正規化画像生成部14bでは、特徴点を基準にして正規化を施す。図4に瞳、鼻穴を基準にした正規化処理の例を示す。ベクトルE1E2の向きを平行に補正してさらにc1, c2の上から1/3の点CPを正規化画像の中心座標、横幅はベクトルE1E2の2倍、縦幅はベクトルc1c2の2倍に設定する。

【0032】パターンマッチング部14cでは、正規化画像と辞書画像に蓄えられている顔画像と比較してパタ

ーン類似度を求める。パターン類似度が基準値より高い場合には、本人であると同定する。基準より小さい場合には、他人とする。パターンマッチングの方法としては、相関法、部分空間法、複合類似度法などが適用可能である。

【0033】(所定動作検出部18) 所定動作検出部18は、使用者がキーボード、ボタンプッシュ、または、マウスの操作などの所定の動作を行っているか否かを検出する。検出結果に基づいて辞書生成部15を制御する。

【0034】具体的には、使用者がキーボード操作、ボタンプッシュ、マウスの操作などの所定の動作を行っていることを検知すると、画像入力手段11に使用者の顔の正面画像が写っている可能性が高いため、辞書登録するのに好適であるために所定の動作を行っている間は辞書生成信号を辞書生成部16に送信する。

【0035】(辞書生成部16) 辞書生成部16は、顔特徴点抽出部14a、フレーム評価部16a、正規化画像生成部14b、学習画像記憶部16b及び主成分分析部16cからなる。ここで顔特徴点抽出部14aと正規化画像生成部14bは、認識部14と共通で使用する。図5に辞書生成部16のブロック図を示す。

【0036】辞書生成は、所定動作検出部18から辞書生成信号が入力されると学習データの収集が行われる。

【0037】フレーム評価部16bでは、抽出された顔特徴点の情報から抽出された顔領域画像02が辞書生成に有効か否かを判断する。ここで3つの評価基準を適用するが、個別あるいは組合せて適用してもよい。

【0038】(i) 予め各特徴点の位置関係を規定しておき、この位置関係が崩れた場合には、学習データとして有効でないと判断する。例えば、両瞳、鼻穴、口端の位置関係などである。

【0039】(ii) 辞書画像17に蓄えられた固有顔画像とパターン類似度を評価にする。類似度が基準より小さい場合には有効でないと判断する。基準を高くする程、学習データの均一性を高めることになる。しかし高過ぎると顔向きなどの変化に対応できなくなる。

【0040】あるいは辞書画像17に蓄えられた固有顔

$$X = a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot x_4 \quad (1)$$

$$Y = b_1 \cdot y_1 + b_2 \cdot y_2 + b_3 \cdot y_3 + b_4 \cdot y_4 \quad (2)$$

$$Res = \sqrt{(X - x')^2 + (Y - y')^2} \quad (3)$$

誤差Resは新しい特徴点の座標誤差を示している。誤差が閾値より大きい場合は、抽出位置に誤りが生じているので、このデータは学習データとして有効でない。

【0047】上記(i)～(iii)の評価基準以外にも顔領域の平均輝度や輝度の分散が基準範囲に入っているか否かで判断することも可能である。

【0048】学習画像記憶部16bは、上記有効である

画像の代わりに、事前に正しい特徴点で正規化された学習画像から生成した辞書画像を用いても良い。

【0041】さらに顔向きまで考慮すると極端に横を向いた顔などを除くことが可能である。この場合には、予め方向1～Nの顔向き毎に準備してある顔辞書パターンとの比較を行って、所定の顔向きの辞書に一致する場合に辞書生成に有効であると判断する。

【0042】(iii) 各フレーム毎に抽出された特徴点の関係を調べて目を閉じた学習データを検出する。新しく抽出された特徴点の座標は、既にその前のフレーム1～4で抽出された特徴点の座標の線形結合で表現できる。したがって全てのフレームで正しく特徴点が抽出されていれば、線形結合した際の誤差が小さくなる。逆に目を閉じて眉と瞳を間違えた場合などには、誤差が大きくなる。

【0043】これについて詳細に図6を用いて説明する。

【0044】正射影モデルを仮定すると、3次元モデルを持たずに、任意の方向から見た顔特徴点の2次元座標を、式(1)(2)に示すように4枚の画像上で対応付けられた特徴点の2次元座標の線形結合により表現できる[S. Ullman, R. Basri: "Recognition by Linear Combinations of Models, IEEE Trans. PAMI, Vol. 13, No. 10, pp. 992-1006, 1991.][向川康博、中村裕一、大田友一、"2枚の顔写真を用いた任意方向の顔画像の生成"、情報論、Vol. 37, No. 4, pp. 635-644, 1996.]。この性質を利用して動画像列において既に抽出されている4枚のフレームにおける正しい特徴点の2次元座標値 $(x_1, y_1) \sim (x_4, y_4)$ から新しく検出された特徴点の2次元座標値 $(X, Y)$ を検証する。

【0045】4フレームの画像に対する特徴点の2次元座標と新しい座標 $(X, Y)$ から線形結合係数 $a_i, b_i, (i=1, 4)$ を最小自乗法により求める。今度は逆に得られた結合係数から近似座標を $(x', y')$ を計算して $(X, Y)$ との誤差Resを計算する。

【0046】

【数1】

と判断された学習データを蓄える。蓄えられた学習データ数が規定の数に達したら主成分分析部16cに学習画像を送る。

【0049】主成分分析部16cでは、学習画像記憶部16bに蓄えられた画像データに対して主成分分析(KL展開)適用して固有画像を求める。固有値が大きい方から上位N個の固有画像を辞書画像17に蓄える。

【0050】上記実施例では、動画像からリアルタイムで辞書生成を行う例を説明したが、動画像から所定動作検出部からの検出信号が来たフレームだけを一旦メモリに蓄えてから行うことも可能である。この場合は、辞書生成に有効である可能性が高いフレームだけを対象にするために、処理の高速化と安定性が計れる。

【0051】また、上記実施例では、識別にパターン的な手法である部分空間法を適用しているが、構造的な手法に置き換えても良い。

【0052】(本装置10の動作の説明)次に、本実施例の動作を図7に沿って説明する。

【0053】最初のアクセスは、パスワードなどを入力する通常のログインを行う。あるいは、事前に登録した使用者の辞書を用いて自動ログインを行うことも可能である。

【0054】先ず画像入力部11から入力された画像から顔領域抽出部12において顔領域が抽出されて状況認識部13に送られる。また所定動作検出部18では、マウス操作、キーボード操作、ボタン操作などの所定動作検出信号を状況認識部に送る。状況認識部13では、顔領域の時間的な有無に基づいて認識モードと辞書生成モードの切替えを行う。また所定動作検出信号に基づいて学習データを収集するか否かを制御する。

【0055】先ず時刻T0からT1までは、認識モードに設定され、前回の使用者本人であるかを同定する。

【0056】(i) 認識モード

顔領域は、顔特徴点抽出部14aに送られて瞳、鼻穴、口端が検出される。次にこれらの特徴点を基準に正規化が施される。正規化画像は、パターンマッチング部14cで前回使用者の辞書画像との類似度が計算され、類似度が基準値より高い場合に前回使用していた本人と同定されてアクセス可能になる。基準値より小さい場合はアクセスを禁止する。

【0057】次に時刻T1からは、辞書生成モードに設定される。

【0058】(ii) 辞書生成モード

所定動作検出部18が所定動作を検出して所定動作検出信号がONになっている場合に学習データを収集して辞書画像を生成する。連続して入力される画像から抽出された顔領域画像は顔特徴点抽出部14aに送られて瞳、鼻穴、口端が検出される。抽出された特徴点情報はフレーム評価部16aに送られて、例えば式(3)から誤差が計算される。誤差が閾値より小さい場合には学習データとして有効であると判断され学習画像記憶部16bに蓄えられる。学習画像記憶部16bに規定枚数以上の学習データが蓄えられたら学習データは主成分分析部16cに送られて辞書画像が生成される。

【0059】使用者が離席すると即座にスクリーンロックを起動する。再び使用者が着席すると最初の処理に戻る。

【0060】なお、上記説明では所定動作検出部18が所定動作を検出して所定動作検出信号がONになっている場合に辞書画像を生成していたが、さらに、所定動作検出信号がOFFにあった後、一定期間(例えば、1分間)の間は続けて辞書画像を生成してもよい。

【0061】(変更例)次に、情報端末のセキュリティだけではなく、情報端末のサービスにおける実施例について説明する。

【0062】例えば、顔を登録する場合に、顔領域から得られる特徴量だけではなく、情報端末の操作履歴も同時に記憶することを考える。顔の情報と同時に登録しておくことによって、その人間の識別と同時に、端末を操作していたときに、どのような情報を求めていたのかを履歴から抽出できる。

【0063】情報検索を例とすると、使用者は検索式等を入力する作業を行う。個人識別のために必要な情報に加え、その検索式、検索内容の履歴をデータベースに蓄えておく。蓄えられたデータをもとに、使用者の興味ある内容を推定、抽出し、新たな情報検索を行って保持しておく。また、その使用者が使用していた画面の表示状態も前記データの一つとして保持しておく。

【0064】そして使用者が再び、情報端末へのログイン、操作復帰を行った場合、端末に新たに検索した所望と考えられる情報の提供やその使用者に合わせた画面の表示状態を自動的に行うサービスが可能となる。

【0065】これらの履歴情報は顔の登録情報とリンクして蓄えられることにより、いつこのような情報を検索していたかというタイムスタンプの代わりになり、これに基づいて、いつの時点の顔の状態に近いのかを認識して、その当時の履歴情報に基づいた情報を提示しても良い。

【0066】また、顔の登録の際の検証機構として、表情検出、認識を用いれば、人間の感情状態に応じた情報の操作履歴をとることも可能であり、平常時の辞書、笑ったときの辞書、怒っているときの辞書などを生成することもできる。新たに顔認識を行う際に、表情情報に対応した情報提供を行っても良い。

【0067】これらは顔登録の更新と情報操作履歴の更新を継続的に同時に行うことにより、より確かな情報提供が可能になる。

【0068】これらの実施例は情報端末を例に説明したが、例えばTV、電話など家電製品に置き換えても良い。

【0069】TVの場合には、リモコンに超小型カメラを搭載する。所定の動作はリモコンのボタンを押している動作と定義する。電話の場合には、受話器に超小型CCDが搭載して受話器を上げる動作を所定の動作と定義する。

【0070】さらに、本実施例の装置10を現金自動預け払い機(以下、ATMという)に使用してもよい。

【0071】すなわち、使用者の本人確認のためにこの装置10を使用することにより、他人による取り忘れ、現金の横取りなどの不正使用を確実に防止できる。この場合には、画像入力部11を、ATMの操作パネルの近傍に配置して使用者の顔が確実に写るようにする。また、顔画像と同時に口座番号や更新登録時刻を登録しておいてもよい。

【0072】

【発明の効果】以上、本発明によれば、入力された画像中から新規顔画像を抽出して、これが辞書登録に値する顔画像か否かを判断して、有効と判断した入力顔画像のみを登録するために、不要な顔画像の登録を排除できて辞書精度を上げることができる。

【図面の簡単な説明】

【図1】本実施例の装置のブロック図である。

【図2】本人同定モード及び辞書生成モードの時間的な変化を示す図である。

【図3】認識部のブロック図である。

【図4】瞳、鼻穴を基準にした正規化処理の例を示す図である。

【図5】辞書生成部のブロック図である。

【図6】線形結合した際の誤差を使用した説明図である。

【図7】本実施例の動作を説明する図である。

【符号の説明】

- 11 画像入力部
- 12 顔領域抽出部
- 13 状況認識部
- 14 認識部
- 15 辞書生成部
- 16 アクセス制御部
- 17 辞書画像
- 18 所定動作検出部